

## **Remarks**

Claims 1-20 remain pending in this application after entry of this paper. Reconsideration of this application is respectfully requested in light of the following remarks.

### **Rejection of Claims 9-15**

#### **Under 35 U.S.C. § 112**

Claims 9-15 have been rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the enablement requirement. Applicants disagree. The step of obtaining pitch values of between one and five of each speech item is described in the Summary of the Invention on page 3, lines 4 - 26. The step is further exemplified in the Detailed Description of the Preferred Embodiment on pages 21 - 26.

### **Rejection of Claims 1 and 8**

#### **Under 35 U.S.C. § 102(e) Over Coorman**

Claims 1 and 8 have been rejected under 35 U.S.C. 102(e) as being anticipated by U.S. Patent No. 6,665,641 issued to Coorman et al. ("Coorman").

Regarding the rejection of claim 1, independent claim 1 recites a method for converting text to concatenated voice by utilizing a digital voice library and a set of playback rules. The digital voice library includes a plurality of speech items and a corresponding plurality of voice recordings. Each speech item corresponds to at least one available voice recording wherein multiple voice recordings that correspond to a single speech item represent various inflections of that single speech item. The method includes receiving text data, converting the text data into a sequence of speech items in accordance with the digital voice library. The method further comprises determining a syllable count for each speech item in the sequence of speech items. An impact value is determined for each speech item. A desired inflection for each speech item in the sequence of speech items is determined based on the

syllable count, the impact value for the particular speech item, and the set of playback rules. A sequence of voice recordings is determined by determining a voice recording for each speech item based on the desired inflection for the particular speech item and based on the available voice recordings that correspond to the particular speech item. Voice data is generated based on the sequence of voice recordings by concatenating adjacent recordings in the sequence of voice recordings.

Coorman discloses a speech synthesizer that concatenates speech waveforms from a speech database to convert text to speech. A text processor generates a target specification. Through a waveform selection process, a synthesizer selects waveform candidates from the database based on various criteria including pitch, duration, and coarse pitch continuity. The criteria may be implemented by cost functions which determine which diphones are most suitable to join and/or are good matches. The waveform candidates are then concatenated forming the output speech similar to the target specification.

The Examiner references col. 9, ll. 33-44 and col. 12-15 of Coorman to show the step of determining an impact value for each speech item in the sequence of speech items. However, the Examiner interpreted Applicants' impact value as how well the speech item fits in the concatenated speech rather than as a step in a complex process to determine where in a spoken sentence inflection changes take place. Also, the cost functions described in col. 12-15 of Coorman are used to formulate a candidate waveform that is compared to a target waveform. The candidate waveform in Coorman is assigned a value based upon how well the target and candidate waveforms match. The diphone selected to be concatenated is based upon the value that signifies the most well matched waveform. In contrast, in Applicants' claimed invention the impact value of a particular speech item is based upon how descriptive and/or important a word is and is correspondingly assigned a value between 0-255. The impact value in combination with the syllable count for the particular speech item and a set of playback rules is then used to determine the desired inflection for each speech item.

The Examiner also refers to col. 9, ll. 26-37 of Coorman to describe the step of determining a desired inflection for each speech item in the sequence of speech items based on the syllable count and the impact value for the particular speech item and further based on the set of playback rules. However, col. 9, ll. 26-37 merely describes selection of the candidate speech unit, candidate-to-target matching, and assigning the cost value.

Coorman fails to describe or suggest the method for converting text to concatenated voice as described by Applicants' independent claim 1. Specifically, Coorman does not teach determining an impact value for each speech item and the step of determining a desired inflection for each speech item in the sequence of speech items based on the syllable count and the impact value for the particular speech item and further based on the set of playback rules, in combination with other recited limitations.

Accordingly, Applicant believes that independent claim 1 is patentably distinguishable over the cited art.

Regarding the rejection of claim 8, claim 8 is dependent from claim 1 and is also believed to be patentable. Claim 8 is believed to recite additional patentable subject matter. Claim 8 recites the method of claim 1 wherein the pitch value for each speech item is determined by normalizing the impact value for the particular speech item. The desired inflection for each speech item is further based on the pitch value for the particular speech item.

#### **Rejection of Claims 2 and 16**

##### **Under 35 U.S.C. § 103(a) Over Coorman and Jacks**

Claims 2 and 16 have been rejected under 35 U.S.C. 103(a) as being unpatentable over Coorman in view of U.S. Patent No. 4,692,941 issued to Jacks et al. ("Jacks").

Jacks discloses a system and method for real-time text-to-speech synthesis wherein text is first compared to words in an exception dictionary. If the word is not found in the exception dictionary, then the system applies standard pronunciation rules to the text. The text is then converted into phoneme sequences. A synthesizer translates the phoneme sequences into speech segments using a look-up table comprising variable length portions of waveforms. Unvoiced transitions are produced by a sequence of segments which can be concatenated in forward or reverse order to generate different transitions from the same segments. Voiced transitions are produced by interpolating adjacent phonemes.

Regarding the rejection of claim 2, claim 2 recites the method of claim 1 wherein a plurality of the speech items are glue items and a plurality of the speech items are payload items. The desired inflection for a glue item is based on the desired inflection for the surrounding payload items in the sequence of speech items. Further, the desired inflection for the payload items is based on the desired inflection for nearest payload items in the sequence of speech items.

Applicants maintain that Jacks fails to overcome the deficiencies of Coorman to achieve Applicants' claimed invention. Particularly, Jacks fails to disclose or suggest determining an impact value for each speech item and the step of determining a desired inflection for each speech item in the sequence of speech items based on the syllable count and the impact value for the particular speech item and further based on the set of playback rules, in combination with other recited limitations of independent claim 1. Claim 2 depends from claim 1 and is believed to be patentable for reasons described above for claim 1.

Independent claim 16 recites a method for converting text to concatenated voice by utilizing a digital voice library and a set of playback rules. The method comprises determining a syllable count for each speech item in the sequence of speech items. The impact value for each speech item is determined. A pitch value within a range for each speech item is determined by normalizing the impact value for the particular speech item. The desired inflection for each speech item in the sequence of speech items is based on the syllable count,

Reply to Office Action of February 23, 2004

pitch value and based on the set of playback rules. The playback rules dictate that the desired inflection for a glue item is based on the desired inflection for surrounding payload items and that the desired inflection for a payload item is based on the desired inflection for nearest payload items with priority being given to speech items having a greater pitch value such that the desired inflections are determined first for speech items having the greatest pitch value and thereafter are determined for speech items in order of descending pitch. A sequence of voice recordings is determined by determining a voice recording for each speech item based on the desired inflection for the particular speech item and based on the available voice recordings that correspond to the particular speech item. Voice data is generated based on the sequence of voice recordings by concatenating adjacent recordings in the sequence of voice recordings.

The Examiner has acknowledges that Coorman fails to specifically disclose that the digital voice library includes a plurality of speech items including glue items and payload items with respect to the method for determining desired inflection of the glue items and the payload items in accordance with claim 16. Instead, the Examiner refers to col. 4, ll. 48 - 59 of Jacks to teach this feature. Jacks describes a method of breaking a sentence down into clauses. The sentence structure analyzer then compares each word of a clause to a key word dictionary. The sentence structure analyzer applies standard rules of prosody to the sentence. However, Applicants assert that Jacks fails to provide the stated deficiencies in Coorman to achieve Applicant's claimed invention. Jacks does not disclose or suggest determining an impact value for each speech item in the sequence of speech items, determining a pitch value within a range for each speech item in the sequence of speech items by normalizing the impact value for the particular speech item and determining a desired inflection for each speech item in the sequence of speech items based on the syllable count and the pitch value for the particular speech item and further based on the set of playback rules.

Additionally, there is no motivation to combine the teachings of Coorman and Jacks to achieve the invention as set out in independent claim 16. Also, there is no disclosure in the specification of Coorman to describe or suggest a digital voice library including a plurality of speech items including glue items and payload items that would make it clear that

the absent glue items and payload items are necessarily present in the method and system described. As such, claim 16 is believed to be patentable.

### **Rejection of Claims 3-5 and 17-19**

#### **Under 35 U.S.C. § 103(a) Over Coorman, Jacks, and Minowa**

Claims 3-5 and 17-19 have been rejected under 35 U.S.C. 103(a) as being unpatentable over Coorman in view of Jacks and further in view of U.S. Patent No. 6,438,522 issued to Minowa et al. ("Minowa"). Claims 3-5 depend from independent claim 1 and are also believed to be patentable. Claims 17-19 depend from independent claim 16 and are also believed to be patentable. Minowa fails to disclose or suggest determining an impact value and then using the impact value to determine a desired inflection for each speech item in combination with other limitations as recited by independent claims 1 or 16. Thus, Minowa fails to overcome the deficiencies of Coorman and Jacks to achieve Applicants' claimed invention.

### **Rejection of Claim 6**

#### **Under 35 U.S.C. § 103(a) Over Coorman and Gasper**

Claim 6 has been rejected under 35 U.S.C. 103(a) as being unpatentable over Coorman in view of U.S. Patent No. 5,278,943 issued to Gasper et al. ("Gasper"). Claim 6 depends from claim 1 and is also believed to be patentable. Gasper fails to disclose or suggest determining an impact value and then using the impact value to determine a desired inflection for each speech item in combination with other limitations as recited by independent claims 1 or 16. As such, Gasper fails to overcome the deficiencies of Coorman and Jacks to achieve Applicants' claimed invention.

### **Rejection of Claim 7**

#### **Under 35 U.S.C. § 103(a) Over Coorman, Gasper and Jacks**

Reply to Office Action of February 23, 2004

Claim 7 has been rejected under 35 U.S.C. 103(a) as being unpatentable over Coorman in view of Gasper and Jacks. Claim 7 is a dependent claim and is also believed to be patentable.

### **Rejection of Claim 20**

#### **Under 35 U.S.C. § 103(a) Over Coorman, Jacks, Minowa, and Gasper**

Claim 20 has been rejected under 35 U.S.C. 103(a) as being unpatentable over Coorman in view of Jacks, Minowa, and Gasper. Claim 20 is a dependent claim and is also believed to be patentable.

### **Conclusion**

In summary, independent claims 1 and 16 recite subject matter that is not described or suggested by Coorman, Jacks, Minowa, and Gasper individually, or in combination, nor is there any motivation to combine the cited art to achieve the claimed invention. Specifically, the cited art fails to describe or suggest determining an impact value for each speech item in the sequence of speech items and using the impact value to determine a desired inflection for each speech item in combination with other claimed limitations. The

Reply to Office Action of February 23, 2004

remaining claims 2-15, and 17-20 are dependent claims and are also believed to be patentable.

Respectfully submitted,

**Eliot M. Case**

By Jeremy J. Curcuri  
Jeremy J. Curcuri  
Reg. No. 42,454  
Attorney/Agent for Applicant

Date: 5/24/04

**BROOKS KUSHMAN P.C.**  
1000 Town Center, 22nd Floor  
Southfield, MI 48075-1238  
Phone: 248-358-4400  
Fax: 248-358-3351